Nome e cognome: ______________________________ *Laurea a luglio:* ☐

indirizzo email: ______________________________

Statistica – Esercitazione del 18 maggio 2018 – Gruppo A

A1. Per valutare se l'uso del telefono cellulare influisce sui tempi di reazione durante la guida è stato condotto un esperimento.
Due gruppi di studenti, ciascuno dei quali composto da 36 individui selezionati a caso, sono stati sottoposti ad una simulazione di guida. Durante la simulazione, gli individui di uno dei gruppi sentivano musica alla radio mentre quelli dell'altro gruppo erano impegnati in una conversazione al cellulare. Sono stati misurati i tempi di reazione con cui gli individui hanno schiacciato il pedale del freno in seguito all'accensione di una luce rossa. Per il "gruppo radio" il tempo medio di reazione è stato di 533.7 millisecondi con una deviazione standard di 65.3, per il "gruppo cellulare" il tempo medio di reazione è stato 585.2 con deviazione standard di 89.6.
Ipotizzando che i tempi di reazione di ciascun individuo seguano una distribuzione normale la cui varianza non dipende dal gruppo di appartenenza, stabilire se questi dati permettono di affermare che l'uso del cellulare allunghi i tempi di reazione.

(a) Si considerino le ipotesi seguenti:
(i) L'uso del cellulare alla guida allunga i tempi di reazione;
(ii) L'uso del cellulare alla guida non allunga i tempi di reazione.
Stabilire quale delle due debba essere l'ipotesi nulla e quale l'alternativa.

(b) Esprimere il sistema di ipotesi risultante dal punto precedente in termini di uguaglianze e/o disuguaglianze sui parametri delle distribuzioni.

(c) Svolgere il test con un livello di significatività pari al 5%.

(d) Conseguentemente a quanto visto nei punti precedenti, quale conclusione si può trarre?

---

argomentare tutti i passaggi in modo adeguato

(a) Come ipotesi alternativa bisogna considerare quella che vorremmo poter dimostrare, quindi che l'uso del cellulare allunga i tempi di reazione. Di conseguenza:
H_0 : *L'uso del cellulare alla guida non allunga i tempi di reazione*
H_1 : *L'uso del cellulare alla guida allunga i tempi di reazione*
(b) Indicando con μ_R il valore atteso relativo al "gruppo radio" e μ_C quello relativo al "gruppo cellulare", il sistema di ipotesi diventa

$$H_0 : \mu_C - \mu_R \leq 0 \qquad \textit{contro} \qquad H_0 : \mu_C - \mu_R > 0.$$

(c) Si tratta di un test unilaterale sulla differenza tra medie di normali con uguale varianza (incognita). La statistica test è

$$T = \frac{\bar{X}_C - \bar{X}_R}{S_p\sqrt{\frac{1}{n_C} + \frac{1}{n_R}}}$$

che sotto H_0 ha distribuzione T con $n_R + n_C - 2$ gradi di libertà.

Rifiutiamo l'ipotesi nulla per valori alti della statistica test, quindi la regione critica (di rifiuto) è $\{T \geq t_{0.05,70}\}$.

I dati del campione osservato sono:

$$\bar{x}_C = 585.2 \qquad \bar{x}_R = 533.7 \qquad s_C^2 = 89.6^2 = 8028.16 \qquad s_R^2 = 65.3^2 = 4264.09$$

per cui $s_p = \sqrt{\frac{35 \cdot 8028.16 + 35 \cdot 4264.09}{70}} = 78.40$.

Il valore osservato della statistica test $\frac{585.2-533.7}{78.40\sqrt{2/36}} = 2.787$ *è maggiore di* $t_{0.05,70} = 1.667$ *per cui l'ipotesi nulla va RIFIUTATA.*
(d) L'uso del cellulare alla guida allunga i tempi di reazione.

A2. Siano $\hat{\theta}_1$ e $\hat{\theta}_2$ due stimatori di uno stesso parametro incognito θ. Si consideri un nuovo stimatore ottenuto come combinazione lineare

$$\hat{\theta} = \lambda\hat{\theta}_1 + (1-\lambda)\hat{\theta}_2.$$

Nell'ipotesi che i due stimatori $\hat{\theta}_1$ e $\hat{\theta}_2$ siano indipendenti l'uno dall'altro, siano entrambi non distorti e che abbiano varianze note, pari a v_1 e v_2 rispettivamente,

(a) valutare la distorsione di $\hat{\theta}$;

(b) determinare il valore del peso λ che rende minimo l'errore quadratico medio di $\hat{\theta}$, scrivere l'espressione dello stimatore ottenuta in corrispondenza del valore ottimo di λ.

---

argomentare tutti i passaggi in modo adeguato

(a) Dall'ipotesi di non distorsione di $\hat{\theta}_1$ e $\hat{\theta}_2$ si ha che $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$. Di conseguenza il valore atteso di $\hat{\theta}$ è

$$E(\hat{\theta}) = E(\lambda\hat{\theta}_1 + (1-\lambda)\hat{\theta}_2) = \lambda E(\hat{\theta}_1) + (1-\lambda)E(\hat{\theta}_2) = \theta$$

per cui $\hat{\theta}$ è non distorto.
(b) Dal momento che $\hat{\theta}$ è non distorto e che $\hat{\theta}_1$ e $\hat{\theta}_2$ sono indipendenti, si ha che

$$MSE(\hat{\theta}) = V(\hat{\theta}) = V(\lambda\hat{\theta}_1 + (1-\lambda)\hat{\theta}_2) = \lambda^2 V(\hat{\theta}_1) + (1-\lambda)^2 V(\hat{\theta}_2) = \lambda^2 v_1 + (1-\lambda)^2 v_2.$$

Per determinare il λ ottimo, minimizziamo $MSE(\hat{\theta})$ rispetto a λ:

$$\frac{d}{d\lambda} MSE(\hat{\theta}) = 2\lambda v_1 - 2(1-\lambda)v_2 = 2\lambda(v_1+v_2) - 2v_2$$

da cui si vede facilmente che il minimo si ottiene per $\lambda = \frac{v_2}{v_1+v_2}$.
Lo stimatore ottimo è

$$\hat{\theta} = \frac{v_2}{v_1+v_2}\hat{\theta}_1 + \frac{v_1}{v_1+v_2}\hat{\theta}_2.$$

A3. Un indicatore della forza fisica di un individuo è dato dal peso massimo che l'individuo riesce a sollevare (*maximum bench press*, che indichiamo con `maxBP`). Tuttavia misurare la forza in questo modo può essere pericoloso, soprattutto con persone che hanno poca familiarità con le tecniche di sollevamento pesi.
Si è cercato quindi di determinare una variabile alternativa che permetta di dedurre il *maximum bench press* con una buona precisione. Una possibile variabile alternativa, che indichiamo con `BP60`, è costituita dal numero di volte che l'individuo riesce a sollevare 60 libbre (un peso relativamente basso e quindi poco pericoloso) prima di essere troppo stanco.
In uno studio sono state rilevate le due variabili `BP60` e `maxBP` su 52 atlete di una high school degli Stati Uniti. Seguono i valori di sintesi dei dati osservati (indichiamo con v_i i valori relativi a `maxBP` e con w_i quelli relativi a `BP60`):

$$\sum_{i=1}^{52} v_i = 4100 \qquad \sum_{i=1}^{52} v_i^2 = 331350 \qquad \sum_{i=1}^{52} w_i = 565 \qquad \sum_{i=1}^{52} w_i^2 = 8889 \qquad \sum_{i=1}^{52} v_i w_i = 48600$$

Si vuole costruire un modello regressione per dedurre `maxBP` da future osservazioni di `BP60`.

(a) Stabilire in quale delle due variabili è opportuno sia la variabile indipendente x e quale la variabile dipendente y.

(b) Determinare i parametri della regressione.

(c) Un'atleta non appartenente al campione studiato ha un valore `BP60` pari a 13. Costruire un intervallo che con confidenza al 95% contenga il valore corrispondente di `maxBP`.

---

argomentare tutti i passaggi in modo adeguato

(a) La variabile indipendente è quella che sarà osservata, la variabile dipendente quella da prevedere. Quindi $x =$ `BP60` *e* $Y =$ `maxBP`.
(b) Dai dati otteniamo

$$\bar{x} = 10.86538 \quad \bar{y} = 78.84615 \quad S_{xx} = 2750.058 \quad S_{YY} = 8080.769 \quad S_{xY} = 4051.923$$

quindi $B = 1.473395$, $A = 62.83715$ *e la retta di regressione è* $y = 62.83715 + 1.473395x$.
(c) L'intervallo di previsione della Y *relativo ad una osservazione in cui la variabile indipendente assume valore* x_0 *è*

$$A + Bx_0 \mp t_{\alpha/2,n-2}\sqrt{\left[1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right]\frac{SS_R}{n-2}}.$$

Dal momento che $t_{0.025,50} = 2.009$ *e che* $SS_R = S_{YY} - S_{xY}^2/S_{xx} = 2110.684$, *l'intervallo è*

$$62.83715 + 1.473395 \cdot 13 \mp 2.009\sqrt{\left[1 + \frac{1}{52} + \frac{(10.86538 - 13)^2}{2750.058}\right]\frac{2110.684}{50}} = (68.80278, 95.17979).$$